

Developing Improved Algorithms for Detection and Analysis of Skin Cancer

By

Ammara Masood

Submitted in partial fulfilment of the requirements for the

Doctor of Philosophy

Faculty of Engineering and Information Technology
UNIVERSITY OF TECHNOLOGY, SYDNEY

July, 2016

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgment

Over the last four years I have had the privilege of working with many people who have made my time at University of Technology, Sydney an enjoyable and intellectually stimulating experience.

I would like to thank all the people who have helped me along the way and contributed to this thesis. I am especially grateful to my Supervisor Associate Professor Dr. Adel Al-Jumaily. He has been very motivating throughout these years. While working with Dr. Adel, I have learned how to pursue my research with an intellectual rigor and how to critically evaluate my work and improve my research skills.

I would like to take this opportunity to acknowledge Scope Global for sponsoring my PhD through the Endeavour Post Graduate Award.

I would like to acknowledge Engr. Omama Masood for doing the proof reading and Engr. Tariq Adnan and Engr. Bani Kaur for helping in formatting corrections.

I would like to thank my friends and colleges who have encouraged and supported me to accomplish my degree.

Finally, I would like to acknowledge the endless love of my family who have been a constant source of support. Their faith in me and encouragement has helped me throughout my life.

Dedication

This thesis is dedicated to the five most important persons of my life without their support this journey would have been just a dream. Special thanks to my Grandfather M.R.A. Khan, my father M. Masood Alam Khan, my mother Azra Muzammil, my aunt Bushra Muzammil and my very best friend Tariq Adnan for being my inspiration and encouraging me throughout my career.

Abstract

Malignant melanoma is one of the deadliest forms of skin cancer and number of cases showed rapid increase in Europe, America, and Australia over the last few decades. Australia has one of the highest rates of skin cancer in the world, at nearly four times the rates in Canada, the US and the UK. Cancer treatment costs constitute more 7.2% of health system costs. However, a recovery rate of around 95% can be achieved if melanoma is detected at an early stage. Early diagnosis is obviously dependent upon accurate assessment by a medical practitioner. The variations of diagnosis are sufficiency large and there is a lack of detail of the test methods. This thesis investigates the methods for automated analysis of skin images to develop improved algorithms and to extend the functionality of the existing methods used in various stages of the automated diagnostic system. This in the long run can provide an alternative basis for researchers to experiment new and existing methodologies for skin cancer detection and diagnosis to help the medical practitioners.

The objective is to have a detailed investigation for the requirements of automated skin cancer diagnostic systems, improve and develop relevant segmentation, feature

selection and classification methods to deal with complex structures present in both dermoscopic/digital images and histopathological images.

During the course of this thesis, several algorithms were developed. These algorithms were used in skin cancer diagnosis studies and some of them can also be applied in wider machine learning areas. The most important contributions of this thesis can be summarized as below:

- Developing new segmentation algorithms designed specifically for skin cancer images including digital images of lesions and histopathological images with attention to their respective properties. The proposed algorithm uses a two-stage approach. Initially coarse segmentation of lesion area is done based on histogram analysis based orientation sensitive fuzzy C Mean clustering algorithm. The result of stage 1 is used for the initialization of a level set based algorithm developed for detecting finer differentiating details. The proposed algorithms achieved true detection rate of around 93% for external skin lesion images and around 88% for histopathological images.

- Developing adaptive differential evolution based feature selection and parameter optimization algorithm. The proposed method is aimed to come up with an efficient approach to provide good accuracy for the skin cancer detection, while taking care of number of features and parameter tuning of feature selection and classification algorithm, as they all play important role in the overall analysis phase. The proposed method was also tested on 10 standard datasets for different kind of cancers and results shows improved performance for all the datasets compared to various state-of-the-art methods.

- Proposing a parallelized knowledge based learning model which can make better use of the differentiating features along with increasing the generalization capability of the classification phase using advised support vector machine. Two classification algorithms were also developed for skin cancer data analysis, which can make use of both labelled and unlabelled data for training. First one is based on semi advised support vector machine. While the second one based on Deep Learning approach. The method of integrating the results of these two methods is also proposed. The experimental analysis showed very promising results for the appropriate diagnosis of melanoma. The classification accuracy achieved with the help of proposed algorithms was around 95% for external skin lesion classification and around 92 % for histopathological image analysis.

Skin cancer dataset used in this thesis is obtained mainly from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. While for comparative analysis and benchmarking of the few algorithms some standard online cancer datasets were also used. Obtained result shows a good performance in segmentation and classification and can form the basis of more advanced computer aided diagnostic systems. While in future, the developed algorithms can also be extended for other kind of image analysis applications.

Table of Contents

List of Abbreviations.....	xiv
List of Figures	xviii
List of Tables.....	xxiii
Chapter 1	1
Introduction	1
1.1 Problem & Motivation	1
1.2 Thesis Aims and Contributions.....	2
1.2.1 Thesis aim and objectives.....	2
1.2.2 Contributions of the Thesis.....	3
A. Contributions to Diagnosis of Skin Cancer	3
B. Contributions beyond Skin Cancer Case Study	4
1.2.3 Structure of the Thesis	4
1.2.4 Publications Resulting from the Thesis	6
1.2.5 Summary	8
Chapter 2	9
Overview on Skin Cancer and Diagnosis Methods	9
2.1 Introduction.....	9
2.2 Human Skin.....	9
2.2.1 Types of Skin	9
2.2.2 Layers of Human Skin	10
2.3 Cancer	11
2.3.1. Development stages of Cancer disease	12
2.4 Skin Cancer	12
2.4.1 Basal Cell Carcinoma (BCC).....	14
2.4.2 Squamous Cell Carcinoma (SCC)	14

2.5	Melanoma.....	15
2.6	Detection & Diagnosis Methods by Medical Experts.....	16
2.6.1	Skin Lesion Imaging Methods	16
•	2.6.1.1 Traditional Photography.....	16
•	2.6.1.2 Dermoscopy Imaging Techniques.....	16
•	2.6.1.3 Multi Spectral Photography.....	17
•	2.6.1.4 Ultrasound Visualization.....	17
2.6.2	Dermoscopic Diagnosis of Skin Lesions.....	17
2.6.3	Diagnosis Techniques	19
•	2.6.3.1 ABCD-E Rule.....	19
•	2.6.3.2 Pattern Analysis.....	22
•	2.6.3.3 The Seven-Point Checklist	24
•	2.6.3.4 The Menzies method	27
•	2.6.3.5 The Three-Point Checklist.....	29
2.6.4	Diagnosis based on Pathological Analysis	31
•	2.6.4.1 Histo-pathological Images.....	32
•	2.6.4.2 Radial Growth Phase Melanoma	32
•	2.6.4.3 Vertical Growth Phase.....	33
•	2.6.4.4 Histologic type (The microscopic structure) of melanoma	33
•	2.6.4.5 Thickness (Breslow).....	34
•	2.6.4.6 Microscopic Satellites	34
•	2.6.4.7 Blood Vessel and Lymphatic Invasion.....	35
•	2.6.4.8 Pathologic features of a melanoma	36
2.7	Summary	36
	Chapter 3	37
	Methodological Review	37
3.1	Introduction.....	37
3.2	Importance of Computer Aided Diagnosis	37

3.3	Development of Computer-aided diagnosis system.....	39
3.4	Image acquisition/ methods for screening skin lesions.....	40
3.5	Pre-Processing.....	44
3.6	Segmentation.....	47
3.6.1	Discontinuity Based Segmentation.....	48
3.6.2	Similarity Based Segmentation.....	49
• 3.6.2.1	Thresholding Approach.....	49
• 3.6.2.2	Region based segmentation.....	50
3.6.3	Segmentation Using Soft-Computing Approaches.....	51
3.7	Feature extraction.....	53
3.8	Feature selection.....	57
3.8.1	Evaluation Measures.....	59
• 3.8.1.1	Filter Methods.....	60
• 3.8.1.2	Wrapper Methods.....	61
• 3.8.1.3	Embedded Methods.....	62
3.8.2	Search Strategy.....	62
• 3.8.2.1	Exhaustive Search.....	63
• 3.8.2.2	Branch and Bound.....	63
• 3.8.2.3	Sequential search algorithms.....	63
• 3.8.2.4	Evolutionary Computation algorithms.....	64
A.	Genetic algorithm.....	65
B.	Ant Colony Optimization.....	65
C.	Particle Swarm Optimization.....	66
D.	Differential Evolution.....	67
3.9	Feature Normalization.....	68
3.10	Classification.....	68
3.10.1	K-nearest neighbor algorithm.....	70
3.10.2	Decision Trees.....	72
3.10.3	Logistic Regression.....	73

3.10.4	ANN.....	73
3.10.5	Support Vector Machines	77
3.10.6	Evaluation of Classification Performance	81
3.10.7	Selection of suitable classification method.....	83
3.11	Model Validation	90
3.12	Summary	92
Chapter 4	93
Thesis Contributions to Segmentation Methods	93
4.1	Introduction.....	93
4.2	Background of Proposed Method.....	93
4.3	Proposed Approach	96
4.3.1	Image Pre-processing.....	96
4.3.2	Histogram analysis based Fuzzy C Mean thresholding (H-FCM).....	97
A.	Fuzzy C mean Clustering.....	100
B.	Selection of Threshold	101
4.3.3.	Morphological operations	104
4.3.4	H-FCM based fast Level Set Segmentation.....	105
4.4	Segmentation of histopathological images for skin cancer.....	108
4.4.1	Pre-Processing	109
4.4.2	Segmentation	112
• 4.4.2.1	Orientation Sensitive Fuzzy C mean Clustering	112
4.5	Experimental Results for Dermoscopic Images.....	114
4.5.1	Image Dataset	114
4.5.2	Statistical Evaluation Parameters.....	117
4.5.3	Comparative Analysis of segmentation results	119
4.6	Experimental Results for Histopathological Images	126
4.7	Summary	128

Chapter 5	129
Thesis Contributions to Feature Extraction and Selection Methods	129
5.1 Introduction.....	129
5.2 Feature Extraction for Skin Cancer Identification	129
5.2.1 Features based on Physical Shape and Colours	130
• 5.2.1.1 Shape features.....	131
• 5.2.1.2 Colour Features	133
5.2.2 Features based on Statistical Analysis	135
• 5.2.2.1 Intensity Histogram Features.....	136
• 5.2.2.2 GLCM based features.....	136
• 5.2.2.3 GLRLM based features	140
• 5.2.2.4 GTDM based features	142
• 5.2.2.5 Autoregressive Modelling based features	147
5.2.3 Feature extraction based on transform based approach	148
• 5.2.3.1 Feature extraction based on Wavelet transform	149
• 5.2.3.2 Feature extraction based on Gabor wavelet	150
• 5.2.3.3 Fuzzy MI-based Wavelet-Packet Algorithm for Feature Extraction..	151
5.3 Feature Selection.....	153
5.3.1 Background of Proposed Feature Selection Method	154
• 5.3.1.1 Differential Evolution.....	155
5.3.2 Proposed Adaptive Differential Evolution based Feature Selection	157
• 5.3.2.1 Computational Steps of ATDEFS Algorithm	166
5.3.3 Experimental Analysis	167
• 5.3.3.1 Experimental Model	167
• 5.3.3.2 Test Datasets.....	168
• 5.3.3.3 Comparative analysis	169
5.4 Summary	175

Chapter 6	176
Thesis Contributions to Classification Methods	176
6.1. Introduction.....	176
6.2 Classification model based on labeled training data.....	176
6.2.1. Support Vector Machine	178
6.2.2 Advised weighted Support Vector Machine	181
6.2.3 Analysis of proposed classification model and discussion	184
• 6.2.3.1 Calculation of Expertise Weight (EW).....	184
• 6.2.3.2 Analysis of parallelized classification model	187
• 6.2.3.3 Experimental Results for Parallelized Classification Model.....	188
6.3 Semi Supervised learning model for Classification	189
6.3.1 Semi Advised-SVM	191
6.3.2 Deep Belief Network	194
6.3.3 Experimental Results for Semi Supervised Learning Model.....	197
6.4 Summary	202
Chapter 7	204
Conclusion and Future Research.....	204
7.1. Introduction.....	204
7.2 Conclusion	204
7.3 Future Works.....	206
7.4 Summary	207
References	207

List of Abbreviations

Adaptive Differential evolution (ADE)
Adaptive Sequential Forward Floating Selection (ASFFS)
Adaptive Snake (AS)
Adaptive Thresholding based LS (AT-LS)
Advised Weighted Support Vector Machine (A-W-SVM)
Anisotropic Diffusion Filters (ADF)
Ant Colony Optimization (ACO)
Artificial Neural Network (ANN)
Atypical Pigment Network (APN)
Basal Cell Carcinoma (BCC)
Binary Particle Swarm Optimization (BPSO)
Circularity Index (CI)
Computer Aided Diagnosis (CAD)
Confocal Scanning Laser Microscopy (CSLM)
Derivatives of Gaussian (DOG)
Dice Similarity Coefficient (DSC)
Differential Evolution (DE)
Digital Dermoscopy Analysis (DDA)
Discrete Wavelet Transform (DWT)
Epiluminescence Microscopy (ELM)
Expectation Maximization (EM)
Expectation Maximization based level set (EM-LS)

Expertise Weight (EW)

False Negative Error (FNE)

False Positive Error (FPE)

Fast Fourier Transform (FFT)

Genetic Algorithm (GA)

Geodesic active Contour (GAC)

Gradient Vector Flow (GVF)

Gray Level Run length Matrix (GLRLM)

Gray Tone difference Matrix (GTDM)

Grey Level Co-Occurrence Matrix (GLCM)

Ground Truth (GT)

Hammoude Distance (HM)

Histogram analysis based Fuzzy C Mean thresholding (H-FCM)

Hue Saturation Lightness (HSL)

Hue-Saturation- Intensity (HSI)

Hybrid Genetic Algorithm (HGA)

Improved Binary Particle Swarm Optimization (IBPSO)

Independent Histogram Pursuit (IHP)

K Nearest Neighbour (KNN)

K mean clustering based LS (K-LS)

Karhunen-Loève (KL)

Laplacian-of-Gaussian (LoG)

Level Set (LS)

Levenberg–Marquardt (LM)

Logistic regression (LogR)

Magnetic Resonance Imaging (MRI)

Malignant Melanoma (MM)

Multiple Layer Perceptron (MLP)

Optical Coherence Tomography (OCT)

Particle Swarm Optimization (PSO)

Peer Group Filter (PGF)

Pigmented Skin Lesions (PSLs)

Pigmented Spitz Naevi (PSNs)

Principal Component Analysis (PCA)

Radial Basis Function (RBF)

Receiver Operating Characteristic (ROC)

Restricted Boltzmann Machines (RBM)

Scaled Conjugate Gradient (SCG)

Segmented Region (SR)

Sequential Backward Elimination (SBE)

Sequential Floating Forward Selection (SFFS)

Sequential Forward Selection (SFS)

Spatial Fuzzy Clustering based LS (SF-LS)

Squamous Cell Carcinoma (SCC)

Statistical Region Merging (SRM)

Support Vector Machine (SVM)

Support Vectors (SV)

Total Body Photograph (TBP)

Total Dermatoscopy Score (TDS)

Transductive Support Vector Machines (TSVM)

True Positive Rate (TPR)

Typical Pigment Network (TPN)

Wavelet Packet Transform (WPT)

Wavelet Packet Transform (WPT)

List of Figures

Chapter 2

Figure 2.1 Types of Human Skin.....	10
Figure 2.2 Layers of Human Skin	10
Figure 2.3 Structure of the Epidermis	11
Figure 2.4 Stages of Cancer Development and Metastasis.....	12
Figure 2.5 Details in Skin Layers.....	13
Figure 2.6 Basal Cell Carcinoma (BCC).....	14
Figure 2.7 Squamous Cell Carcinoma (SCC).....	14
Figure 2.8 Melanoma	15
Figure 2.9 Traditional Imaging Vs. Dermoscopy.....	17
Figure 2.10 Macroscopic Image Vs. Dermatoscopic Image.....	18
Figure 2.11 ABCD-E rule for Diagnosis of Melanoma.....	21
Figure 2.12 Application of ABCD rule.....	22
Figure 2.13 Melanoma with a blue-white veil	22
Figure 2.14 3 classes of pigment network.....	23
Figure 2.15 Examples of Streaks	24
Figure 2.16 Texture characteristics of skin lesions	26
Figure 2.17 Classification using 7-point checklist	27
Figure 2.18 Application of Menzies method to melanocytic lesions	28
Figure 2.19 Diagnosis of Melanoma using Three-point Checklist.....	30

Figure 2.20 Manual Segmentation by Pathologist.....	32
Figure 2.21 Radial Growth Phase of Melanoma	33
Figure 2.22 Vertical growth Phase of Melanoma.....	34
Figure 2.23 Microscopic Satellites.....	35
Figure 2.24 Blood Vessel and Lymphatic Invasion.....	36

Chapter 3

Figure 3.1 Computer Aided Diagnostic System for Skin Cancer Diagnosis.....	40
Figure 3.2 Categorization of Image Segmentation Algorithms.....	48
Figure 3.3 Categorization of Similarity based Segmentation Algorithms	49
Figure 3.4 Categorization of Region based Segmentation Algorithms.....	50
Figure 3.5 Illustration of feature distribution used in dermoscopic studies in literature.....	56
Figure 3.6 Working Mechanism of Artificial Neural Network.....	74
Figure 3.7 Principle of Support Vector Machine	77
Figure 3.8 Illustration of Classification methods as used by existing diagnostic systems in literature	85

Chapter 4

Figure 4.1 Histogram Analysis based Fuzzy C mean Thresholding.....	100
Figure 4.2 Thresholding of image with melanoma lesion.....	103
Figure 4.3 Thresholding of image with melanoma lesion.....	104
Figure 4.4 Thresholding of image with benign lesion	104
Figure 4.5 Flowchart of proposed algorithm	108

Figure 4.6 Pre-processing stage for Histopathological images.....	111
Figure 4.7 Proposed segmentation method for histopathological images	112
Figure 4.8 Difficulties in skin lesion images	115
Figure 4.9 Segmentation of images with uneven and abrupt lesion border	115
Figure 4.10 Segmentation of skin lesion surrounded by hair.....	115
Figure 4.11 Segmentation of lesion with multiple colours	116
Figure 4.12 Segmentation of lesions with poor illumination.....	116
Figure 4.13 Segmentation of image having specular reflection... ..	116
Figure 4.14 Segmentation of lesion with prominent skin lines	116
Figure 4.15 Segmentation of image with smooth transaction between lesion and skin.....	116
Figure 4.16 Segmentation of image with bubbles of dermoscopic gel.....	116
Figure 4.17 Segmentation of image with multiple lesions.....	116
Figure 4.18 Diagrammatical representation of evaluation parameters	118
Figure 4.19 Border detection results obtained with (a) Ground truth by expert (b) Region based active contour (c) Adaptive thresholding (d) Fuzzy c-mean clustering (e) K-mean clustering (f) Proposed algorithm.....	120
Figure 4.20 Border detection results obtained with (a) Ground truth by expert (b) Region based active contour (c) Adaptive thresholding (d) Fuzzy c-mean clustering (e) K-mean clustering (f) Proposed algorithm.....	120
Figure 4.21 Segmentation results of a melanoma with hair and uneven border.....	123
Figure 4.22 Segmentation results of a dysplastic nevi with smooth transaction between skin and lesion.....	123
Figure 4.23 Segmentation results of a melanoma present on spotty skin with redness effect around lesion	123

Figure 4.24 Segmentation results of a benign lesion surrounded by hair.....	124
Figure 4.25 Segmentation results of a melanoma with multiple colours and prominent skin lines.....	124
Figure 4.26 Segmentation results of a dysplastic nevi with specular reflection and uneven illumination.....	124
Figure 4.27 Segmentation results of the proposed method	127

Chapter 5

Figure 5.1 Experimental Learning Model.....	168
Figure 5.2 Average Classification Accuracies across 20 runs for each subset size across different datasets	173

Chapter 6

Figure 6.1 Proposed Computer Aided Diagnostic Model	178
Figure 6.2 Basic ideas of support vector machines	179
Figure 6.3 Advised Weight Support Vector Machine	183
Figure 6.4 Feature extraction sets used in experiments	184
Figure 6.5 Diagnosis expertise analysis of suggested feature sets.....	185
Figure 6.6 Parallelized Classification model	187
Figure 6.7 Diagnosis results based on Expertise weight	187
Figure 6.8 10-fold cross validation results for the proposed model.....	188
Figure 6.9 Ensemble Learning Model	191
Figure 6.10 Confidence score distribution Vs total correct decisions % for melanoma lesion identification.....	198

Figure 6.11 Confidence score distribution Vs total correct decisions % for benign lesion identification.....	199
Figure 6.12 Confidence score distribution Vs total correct decisions % for cancerous region identification in histopathological images	199
Figure 6.13 Performance comparison of semi advised SVM for dermoscopic and Histopathological image	201
Figure 6.14 Performance comparison of learning model proposed for histopathological image classification	202

List of Tables

Chapter 2

Table 2.1 Calculation of ABCD score of Dermatoscopy.....	20
Table 2.2 Interpretation of TDS Results.....	21
Table 2.3 Seven-Point Checklist for the dermoscopic differentiation between benign melanocytic lesions and melanoma.....	25
Table 2.4 Menzies Method.....	28
Table 2.5 Definition of Three Point Checklist.....	29

Chapter 3

Table 3.1 In Vivo Imaging Techniques for Diagnosis of Skin Cancer	41
Table 3.2 Methods for Segmentation of Dermoscopic Images.....	52
Table 3.3 Algorithm output of KNN according to the area size	71
Table 3.4 Measures for Evaluating Performance of a Classifier.....	82
Table 3.5 Classification Methods used in Literature for Skin Cancer Diagnosis	84
Table 3.6 Summary of Classification Performance of some skin Cancer Detection Methods Present in Literature	86
Table 3.7 Percentage of Diagnostic Models Satisfying the Quality Assessment	91

Chapter 4

Table 4.1 Comparison results of segmentation methods.	121
Table 4.2 Results of Segmentation Methods.	126

Table 4.3 Comparative Analysis of Segmentation for Histopathological Images..	128
---	-----

Chapter 5

Table 5.1 GLCM based Features.....	138
Table 5.2 GLRLM based Features.....	141
Table 5.3 GTDM based Features.....	147
Table 5.4 Description of Datasets used in experiments.....	169

Chapter 6

Table 6.1 Comparative Analysis of Classification results.....	189
---	-----